Resource allocation planning for CERN in 2006

Bernd Panzer-Steindel, CERN/IT

Draft v1 11.01.2006
Draft v2 26.01.2006
Draft v3 03.02.2006
Draft v4 06.03.2006

The note gives an overview of the current status for the planning and allocation of CPU, disk and tape storage resources in 2006.
It is a 'working' document and will be updated in regular intervals.

The table shows the planned allocation of the different resources per experiment in 2006 :

Table 1

Resources 2006	**CPU [KSI2000] End Mar (Sept)**	**Disk space [TB] End Mar**	(additional)**Tape space [PB] End Mar (Summer)**
ALICE	693 (1040)	231	0.17 (0.25)
ATLAS	367 (550)	176	0.33 (0.5)
CMS	1080 (1620)	176	0.33 (0.5)
LHCb	140 (210)	188	0.2 (0.3)
SUM	**2280 (3420)**	**771**	**1.0 (1.55)**

The given number are calculated taking into account the following points:

1. the LHCb numbers are from their TDR
2. the ALICE, ATLAS and CMS numbers are based on a linear extrapolation between the average resource usage numbers from 2005 and the TDR numbers in 2007
3. there is no distinction between T0 and CAF, the numbers represent the sum of the two values.
4. Due to budget constraints in 2006 a cut of ~50% was applied to the CAF numbers which corresponds to a 10% overall cut

5. in general the availability of the resources for the experiments in 2006 was planned to be : 2/3 at the end of February and another 1/3 in September

There are a few details to be considered for the three different resources.

CPU

The following table shows the CPU resource requested in 2005 and how much was actually used in 2005. The distribution is not homogeneous over the year due to the different productions and data challenges. The numbers do not contain the extra usage from the ATLAS DAQ challenge in June 2005 (up to 700 nodes occupied).

Table 2

CPU 2005	**Requested [KSI2000]**	**Average [KSI2000]**	**Highest four week Average [KSI2000]**	**Highest one week Average [KSI2000]**
ALICE	100	27	87	110
ATLAS	200	224	560	711
CMS	400	323	563	587
LHCb	40	106	449	482

Today there are about 1150 nodes with a capacity of ~2000 KSI2000 available in LXBATCH. We are currently installing another 1200 nodes with a capacity of ~2400 KSI2000 which will be available at the end of February.
The sum of this has to be split into the following parts :

- ~540 KSI2000 for the dedicated data challenge setup (= 300 nodes)
- ~850 KSI2000 for the fixed target experiments plus AB (= 500 nodes)
- ~510 KSI2000 for replacement of old GRID testbed nodes (= 300 nodes)
- ~2300 KSI2000 for the 4 LHC experiments (= 1150 nodes)

The next delivery of CPU nodes will take place during late summer and will add in September 400 nodes with a capacity of ~1100 KSI2000.
All the numbers have an error-bar of about 5%.

Disk storage

The distribution of disk space in 2005 was about 25 TB less than the aggregate request from the 4 experiments (allocation: ALICE = 38 TB, ATLAS = 28 TB, CMS = 29 TB, LHCb = 13 TB). This was mainly due to the scalability problems with Castor1 where the

stager implementation was not always able to cope with the large number of files in the disk pools. Thus adding more disk space would have made the situation worse.

The new disk space will be used under Castor2 and the migration of the experiments has already started. With the current deliveries of new disk space we most probably have already enough capacity to give 100% of the anticipated disk storage to all 4 experiments by the end of February. The given numbers include all existing disk space, i.e. 770 TB is the total amount of disk space available for the 4 experiments.

Tape storage

This table shows the current (end January 2006) amount of tape space occupied for the different experiments with the highest usage at CERN :

Table 3

Tape storage	Tape storage [TB]	Number of files [million]
COMPASS	1178	5.9
NA48	651	4.6
nTof	258	0.5
ALICE	292	2.8
ATLAS	209	4.6
CMS	308	3.0
LHCb	161	1.4
Total Castor	**4380**	**45.8**

The situation with free tape space is relatively complicated for the moment. We have more than 1 PB still free on the production 9940 STK tapes, so there is no immediate problem. But we have of course also to consider the fixed target experiments and other user communities which are estimated to use about 0.8 PB of extra tape storage in 2006. In addition we are in the middle of a tape evaluation exercise with IBM and STK where we have to decide in summer which is the next generation of tape storage to be used from 2007 onwards. On the other side we have to consider that buying and writing to new 9940B tapes in 2006 is not unproblematic, because we have to essentially copy these data already in 2007 (or end 2006) to the new tape storage.
The details here have probably to be discussed with each experiment separately and depend on whether data is created in data challenges or productions, i.e. where do we need already long term storage and where can we over-write data.

Activities in 2006

The listed tests and productions can be scheduled on two different facilities at CERN. We have the production system available with the resources described in the previous

chapters. In addition a dedicated DRC (Data Recording Challenge) test facility can be used (~300 CPU nodes, 48 Disk server and 40 tape drives) for scalability and performance tests in a large scale. This will primarily be used for the IT tests and coordinated experiment activities (e.g. ALICE DAQ test). The resources can also be moved for short time periods into the production system to overcome resource bottlenecks.

IT

1) **February** tape storage at 1.0 GB/s
2) **March** T0 data recording at 0.5 GB/s
3) **April** T0 data recording at 1.0 GB/s
4) **April** SC4 throughput tests
5) **May** tape storage at 2.0 GB/s
6) **May-August** T0 scalability tests, reconstruction, calibration, analysis scenarios
7) **September** data recording at 1.6 GB/s

The different T0 exercises for IT need up to 50 disk server , 300 CPU nodes and 40 tape drives.

ALICE

1) **March-April** T0-T1 “loop-back” test, MC production on T1+T2, RAW+ESD back to CERN, 100M pp events + 1M PbPb events, 243 TB of data stored at CERN, 231.5 MSI2000 days = 7500 CPUs for 31 days (1KSI2000 per CPU) 31 days == 110 MB/s data rate to CERN
2) **April** T0-T1 disk-disk and disk-tape transfers, RAW to ESD reconstruction on the T1 (FTD-FTS system), CERN only as data export ~ 30 MB/s
3) **May** ALICE DC VII, DAQ-Castor2 at 1.0 GB/s milestone (next ALICE DAQ DC only in 2008), 40 disk server and 25 tape drives
4) **May** test of the T0 facility included in the DAQ-Castor2 test, xRootd access to the data in the T0 buffer (reconstruction, calibration, export functionally and some scalability, 40 disk server, 25 tape drives and 200 CPU nodes
5) **July** T0 to T1 data export (RAW+ESD) with FTS, total data rate from CERN ~400 MB/s ??? (not clear)
6) **September** PROOF analysis facility tests at CERN, ~4.8 Gbyte/s from 26 TB of ESD data, 72000 jobs/day, 1000 CPUs

ATLAS

1) **February** Monte-Carlo production, Grid with CERN share
2) **March-April** 1 week T0 tests
3) **April-May** distributed operation (small testbed) GRID
4) **April-December** Monte-Carlo production
5) **June** 3 weeks T0 tests with Tier-1 export
6) **July** 3 weeks distributed processing tests
7) **July-August** 2 weeks distributed analysis tests
8) **September-October** 3-4 weeks T0 tests with Tier2 centers involved
9) **October** 3 weeks distributed processing tests
10) **October-November** DAQ scalability tests
11) **November** 3-4 weeks distributed analysis test

All T0 exercises for ATLAS need about 40 disk server , 200 CPU nodes and 25 tape drives.

CMS

1) **May-July** T0 tests and preparations
2) **April** export to T1 of 10 TB samples, 150 MB/s
   200 MB/s from the cache in front of MSS at T1 (CERN)
3) **May** production, 10 M events
4) **July-September** production, 25 M events per month, 1 TB per day, storage at CERN for CSA2006
5) **15.September-15.November** CSA2006 (Computing Software and Analysis System test)
6) **November–December** test T0-T1 transfers at aggregate 300 MB/s

LHCb

1) **March-May** Monte-Carlo production, CERN and T1s, 300 TB data at CERN, ~7.8 MSI2000 months needed, 200 M MC events
2) **June-July** Reconstruction phase, ~0.4 MSI2000 months needed, CERN and T1s; stripping exercise in parallel, 0.1 MSI2000 months
3) **October** alignment and calibration challenge, not CPU or storage intensive, requires COOL/3d setup, CERN and T1s

If in the different productions/challenges no specific resources requirements are mentioned, than it is assumed that the given (first chapter) resources at CERN are sufficient.

General

1) SC4 activities
2) PROOF farm (expect soon a request to the MB)

Details about the exact need for resources for the different tests are currently being discussed and will hopefully be clarified during March, as the first clash of resource requests will be already in April.

The most problematic test is the large scale DAQ setup for ATLAS during 4 weeks in October – November. The CMS CSA2006 is ending only in November, so until it ends CMS will not be able to provide additional resource to the planned Atlas DAQ challenge. Either the required additional resources for Atlas could be satisfied from the other two experiments, or the Atlas challenge would need to be pushed back in time.
Including the DRC testbed we will have in total about 1900 nodes available. A possible schedule could look like the following, assuming that the test would start in the second half of November.

Table 4

	# nodes ATLAS DAQ test	**Remaining CPU capacity available for ATLAS**	**Remaining CPU capacity available for ALICE, CMS, LHCb**
Week 44	400	64 %	100 %
Week 45	600	20 %	94 %
Week 46	800	10 %	82 %
Week 47	1200	5 %	55 %